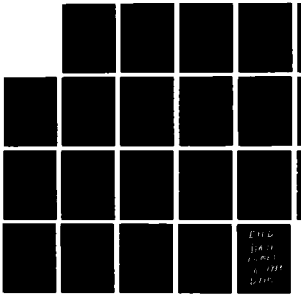
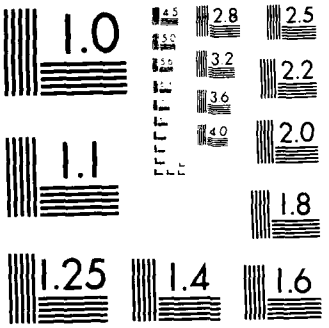
AD-A192 854 ROBUST METHODS FOR ARRAY SIGNAL PROCESSING(U) LOWELL UNIV MA DEPT OF MATHEMATICS C M BYRNE 28 DEC 87 1/1
N00014-87-K-0394

UNCLASSIFIED F/G 17/1 NL





(4)

AD-A192 854

To: Program Manager
Underwater Acoustics
ONR Code 425UA
800 N. Quincy St.
Arlington VA 22217-5000

28 Dec1987

From: Dr. Charles L. Byrne
Chairman, Dept. of Mathematics
Univ. of Lowell, Lowell, MA 01854

Subject: Final Report (ONR Contract N00014-87-K-0394)

The attached final report concerns the work done under the title "Robust methods for array signal processing" at the University of Lowell during the period 15 May 1987 to 30 September 1987.

Robust methods for array signal processing (Final report)

by

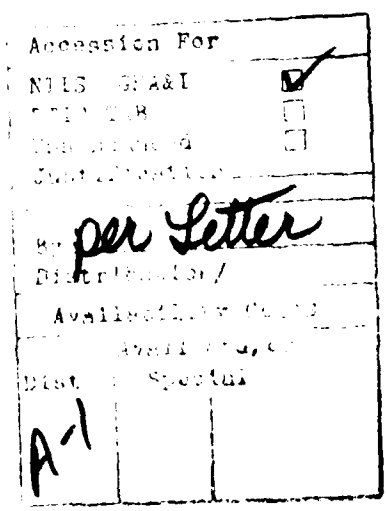
Dr. Charles L. Byrne, Principal Investigator
Department of Mathematics
University of Lowell
Lowell, Massachusetts 01854
(617) 452-5000 (ext. 2520)

ONR Contract N00014-87-K-0394

awarded to

The University of Lowell, Mass.

Contents



1. Summary of work performed

0) Overview: during the period of this contract research was ongoing in several areas: stable array processing (with A. Steele); two-dimensional reconstruction algorithms (with M. Fiddy); extensions of the concept of cross-entropy distance measures (with L. Jones); and the beginning of new work on array processing in more complicated acoustic environments (with NORDA personnel). Other work, by Jones and by J. Benedetto, was partially supported by this contract (title pages enclosed).

1) The array processing work (with A. Steele) In collaboration with Dr. A. Steele I have developed a "sector- focused stability" method for stabilizing nonlinear methods for localization and resolution. This work was performed at DRCS, Salisbury, South Australia (with partial support from the DOD of Australia) and at the University of Lowell, under subcontract. During the period of the present contract we considered the destabilizing effects of short averaging times; a paper on this is in preparation

2) Two-dimensional reconstruction (with M. Fiddy) We considered the problem of reconstruction of two-dimensional distribution functions from spectral values as a problem of obtaining finite-parameter approximations to optimal Wiener filters. The first paper on this subject has just been accepted by Inverse Problems (galleys enclosed) and a second, presenting illustrations, is in preparation.

3) Extension of cross-entropy distances (with L, Jones) Work of L. Jones on extension of cross-entropy distance measures has been applied to obtain new algorithms for incorporating a positivity constraint in nonlinear reconstruction. A paper on this subject is in preparation (a first draft is enclosed).

4) Array processing in a complex acoustic environment (with C. Feuillade, D. DelBalzo et al, NORDA) Standard linear and nonlinear array processing methods cannot be applied unaltered for range, depth and bearing estimation in shallow water situations (with, for example, a normal mode model for the propagation). We have begun to develop new methods for incorporating propagation models in linear and nonlinear array processing. C. Byrne spent a week at NORDA in June 87, supported partly by NORDA; a paper on this work is in progress (first draft enclosed).

2. Background

A previous contract with ONR, for three years and through Catholic University, was terminated at the end of two years rather than continue the subcontract arrangement to Univ. of Lowell. The present contract supported the wrapping up phase of the earlier work with A. Steele and with M. Fiddy, as well as the beginning of new efforts with L. Jones and with NORDA personnel (C. Feuillade and D. DelBalzo). In addition, John Benedetto was supported for two weeks in his investigation of the use of wavelets and Gabor transforms to represent transient signals.

Work supported by ONR Contract N00014-K-87-0394 (Period: May 87-Sept87)

Papers accepted for publication:

1. "Images as power specta; reconstruction as Wiener filter approximation," in Inverse Problems (with M. Fiddy);

Papers in preparation:

1. "Stabilizing eigenvector methods of source localization and resolution for the case of white noise and short averaging time," (with A. Steele);
2. "On entropy critieria for solving inverse problems with positivity constraints," (with L. Jones);
3. "Stable data adaptive matched field methods for ambiguity reduction in source parameter estimation," (with D. DelBalzo and C. Feuillade);
4. "Reconstruction as filter function approximation: some algorithms," (with M. Fiddy).

Papers by other authors receiving partial support from this contract:

1. "Approximation-theoretic derivation of logarithmic entropy principles for inverse problems and unique extension of the maximum entropy principle to incorporate prior knowledge," by Lee Jones (accepted by SIAM J. Applied Math.);
2. "Gabor representations and wavelets," by John Benedetto.

Talks:

1. "Sector-focused stability for high resolution array processing," IEEE Workshop on Underwater Acoustic Signal Processing, Univ. of Rhode Island, Sept. 87 (with A. Steele).

Images as power spectra; reconstruction as a Wiener filter approximation

Charles L Byrne† and Michael A Fiddy‡

† Department of Mathematics, University of Lowell, Lowell MA 01854, USA
‡ Department of Electrical Engineering, University of Lowell, Lowell MA 01854, USA

Received 10 August 1987, in final form 17 September 1987

Abstract. The problem of reconstructing a non-negative function from finitely many values of its Fourier transform is a problem of approximating one function by another and, as such, is analogous to the design of finite-impulse-response approximations to the Wiener filter. Using this analogy we obtain reconstruction methods that are computationally simpler approximations of entropy-based procedures. Our linear estimators allow for the inclusion of prior information about oversampling rate, i.e. support information, as well as other prior knowledge of the general shape of the object. Our nonlinear methods, designed to recover spiky objects, make use of prior information about non-uniformity in the background to avoid bias in the estimation of peak locations.

1. Introduction

The problem of reconstructing a non-negative function $f(a, b)$ of two real variables from finitely many values of its Fourier transform (FT) arises in a number of applications. These include recovering an image or object distribution from its spectrum, a power spectrum from its autocorrelation function, a distribution of energy in bearing from cross-sensor correlations or a bivariate probability density from its characteristic function. In many cases of interest the function $f(a, b)$ is non-negative and we shall make that assumption here. The problem of limited data can arise for a variety of reasons: to remove the effects of a known convolution-filter degradation one can divide by the filter transfer function in the spectral domain, but must avoid dividing by small quantities; in the case of sensor array processing one is limited to spatial separations provided by the array geometry.

Because the data are finite there will always be infinitely many reconstructions consistent with the data values. Some of these reconstructions will be reasonably good, while others will not; the data constraints, by themselves, will not automatically lead to a good reconstruction unless the number of data values is large. There are several methods based on minimising some cost function, such as entropy; one problem with such approaches is that it is not always clear just how the resulting reconstruction is related to the correct answer. The methods we present here are based on the theory of best approximation in Hilbert space and make clear how the reconstruction is related to the original, unknown, correct object function.

In order to obtain a good reconstruction it is necessary to incorporate additional information about the function being reconstructed; in some cases support information is used or positivity is enforced; in others upper and lower bounds are employed. In a number of methods one uses a prior estimate of $f(a, b)$; this is done in cross-entropy minimisation and it has been shown that the Burg maximum entropy method employs (tacitly) a uniform prior estimate [1]. In earlier work [1] work we extended the Burg method for the one-dimensional case to incorporate other prior information and considered numerical examples; our purpose here is to provide a theoretical justification for that procedure, based on analogy with the design of approximate Wiener filters. This approach allows for generalisation to higher dimensions, which we also consider. We present both linear and nonlinear methods.

One of the difficulties with methods that incorporate prior knowledge is that it is not always clear what the prior estimate is estimating. As our development here reveals, the role of the prior estimate is different in the linear and the nonlinear methods. The chief virtue of the Wiener filter design approach is that it gives us a clear picture of the role being played by the prior estimate. Loosely speaking, in the case of linear methods the prior estimate is an estimate of the whole function associated with the data, including any noise background component, while for nonlinear methods (to be used mainly for high-resolution reconstruction of spiky objects) the prior should estimate the smooth component only; linear methods such as superresolution become unstable when the prior estimates only the support-limited object and ignores any noise background, while nonlinear methods that employ a uniform prior estimate, such as Burg's maximum entropy [1], become unstable when the background is non-uniform.

In reference [1] we presented methods for the reconstruction of 1D objects from limited FT data. Here we extend these methods to 2D objects and present a unified interpretation of both cases in terms of the finite-impulse-response approximation to a Wiener filter; in this way we are led naturally to the particular Hilbert spaces used earlier [1], where they may have seemed somewhat *ad hoc*.

It is important to note that, while the Wiener filter and its finite-impulse-response approximations are used to motivate the reconstruction methods presented here, we do not employ a statistical model for the functions being reconstructed.

Throughout the paper we denote by $f(a, b)$ a non-negative function supported on the square $|a| \leq \pi$, $|b| \leq \pi$. The Fourier series representation for the function f on $|a|$, $|b| \leq \pi$ is

$$f(a, b) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} F(m, n) \exp(ima + inb) \tag{1.1}$$

We assume that we have the data $F(m, n)$ for $|m| \leq M$, $|n| \leq N$, from which we are to reconstruct (estimate) $f(a, b)$. A commonly used estimate is the truncated Fourier series (also sometimes referred to as the 'discrete Fourier transform' because the summation replaces the integration); for $|a|$, $|b| \leq \pi$ define the DFT(a, b) to be

$$\text{DFT}(a, b) = \sum_{-M}^{M} \sum_{-N}^{N} F(m, n) \exp(ima + inb). \tag{1.2}$$

Note that the DFT is defined here as a function of two continuous variables; one sometimes sees 'DFT' used to denote a sampled version of (1.2).

In many applications the DFT will be unsatisfactory, particularly if the function f is supported on a smaller interval within $[-\pi, \pi] \times [-\pi, \pi]$, or if f is a spiky function and the number of data values is not large. The DFT is consistent with the original data, in the sense that the Fourier series of DFT (a, b) has the data values in the proper positions, but may fail to be non-negative or to resolve closely spaced peaks. The objective of high-resolution processing is to employ prior information to obtain better reconstructions than the DFT.

For completeness we discuss the Wiener filter and its approximations, for the 1D case (for notational simplicity), and then discuss the use of Wiener filter approximation for the reconstruction of 1D functions. We then turn to the 2D case, the main differences stemming from difficulties in extending the concept of 'causal filter'. Finally, we discuss briefly the connections between these methods and those based on the minimisation of cross-entropy.

2. Wiener filtering: the one-dimensional case

The Wiener filter [2] is a procedure designed to produce as output an estimate of 'signal' when presented with input 'signal plus noise'. Assume that $\{s(n)\}$, $\{u(n)\}$ are independent, mean-zero stationary random sequences with autocorrelation functions $r_{ss}(m)$, $r_{uu}(m)$ and power spectra $R_{ss}(\alpha)$, $R_{uu}(\alpha)$, respectively, with $|\alpha| \leq \pi$; the sequence $\{r_{ss}(m)\}$ are the Fourier coefficients of $R_{ss}(\alpha)$, and similarly for $R_{uu}(\alpha)$. The Wiener filter is a doubly infinite sequence $\{h(k)\}$ designed as follows: given the random sequence $x(n) = s(n) + u(n)$ as input and $y(n)$ as output, where

$$y(n) = \sum_{-\infty}^{\infty} h(k)\, x(n-k) \qquad -\infty < n < \infty \tag{2.1}$$

select $\{h(k)\}$ so as to minimise the expected mean square error, $E|s(n) - y(n)|^2$. The well known result is that the optimal choice of sequence $\{h(k)\}$ is the sequence of Fourier coefficients of the function $H(\alpha) = R_{ss}(\alpha)/R_{xx}(\alpha)$, where $R_{xx}(\alpha) = R_{ss}(\alpha) + R_{uu}(\alpha)$, and $H(\alpha)$ is defined to be zero if $R_{xx}(\alpha) = 0$.

The Wiener filter is not a causal filter, since we do not have $h(k) = 0$ for $k < 0$. We can ask for the causal filter $\{g(k)\}$ $(g(k) = 0, k < 0)$ that best approximates the Wiener filter, or, going further, the finite-impulse-response filter $\{d(k)\}$ $(d(k) = 0$ unless $K \leq k \leq L)$ that best approximates the Wiener filter. To obtain these optimal approximations we minimise the expected mean square difference between the outputs of the Wiener filter and the approximation. These optimisation problems are equivalent to best approximations in a Hilbert space with weighted inner product.

To obtain the best causal approximation to the Wiener filter we minimise the distance

$$\int_{-\pi}^{\pi} \left| H(\alpha) - \sum_{k=0}^{\infty} g(k) \exp(ik\alpha) \right|^2 R_{xx}(\alpha)\, d\alpha \tag{2.2}$$

over all causal sequences $\{g(k)\}$. Similarly, to obtain the optimal finite-impulse-response filter with support $K \leq k \leq L$ we minimize the error

$$\int_{-\pi}^{\pi} \left| H(\alpha) - \sum_{k=K}^{L} d(k) \exp(ik\alpha) \right|^2 R_{xx}(\alpha)\, d\alpha \tag{2.3}$$

over all finite sequences $\{d(k)\}$. From the orthogonality principle in Hilbert space [3] it follows that the optimal $\{g(k)\}$ and $\{d(k)\}$ must satisfy the following systems of linear equations:

$$r_{xx}(m) = \sum_{k=0}^{\infty} g(k)\, r_{xz}(m-k) \qquad m \geq 0 \tag{2.4}$$

$$r_{xz}(m) = \sum_{k=K}^{L} d(k)\, r_{zz}(m-k) \qquad k \leq m \leq L \tag{2.5}$$

Equations (2.4) are the discrete Wiener–Hopf equations. Having solved these equations we write, for $|\alpha| \leq \pi$,

$$G(\alpha) = \sum_{k=0}^{\infty} g(k) \exp(ik\alpha) \qquad D_K^L(\alpha) = \sum_{k=K}^{L} d(k) \exp(ik\alpha) \tag{2.6}$$

It is worth noting that the best finite impulse-response filter approximating $\{g(k)\}$, for $0 \leq K \leq k \leq L$, is the $\{d(k)\}$ above; that is, this choice of $d(k)$ minimises the approximation error

$$\int_{-\pi}^{\pi} \left| G(\alpha) - \sum_{k=K}^{L} d(k) \exp(ik\alpha) \right|^2 d\alpha \tag{2.7}$$

viewed as a function of the $d(k)$. Therefore, the function $D_K^L(\alpha)$ is simultaneously the best approximation, of its form, of $H(\alpha)$ and of $G(\alpha)$, in the Hilbert space with inner product weighted by $R_{xx}(\alpha)$.

The error (2.3) is of interest for the reconstruction problem because, for the case $0 \leq K$, a non-negative function ($H(\alpha)$) is being approximated by a necessarily non-real trigonometric polynomial, in a Hilbert space with weighted inner product. As shown in reference [1], this is precisely what happens in the maximum entropy method (MEM) of Burg [4], that the finite polynomial also approximates $G(\alpha)$ is implicit in the MEM in the spectral factorisation [5].

In the next section we employ these approximation theoretic aspects of Wiener filter design to obtain reconstruction methods.

3. Wiener filter approximation and reconstruction: 1D case

We consider the problem of reconstructing the non-negative function $f(a)$, $|a| \leq \pi$, from finitely many values of its Fourier coefficients, $F(m)$, $|m| \leq M$. We present first linear methods and then nonlinear ones.

3.1. Linear methods

Assume that we have a prior estimate of the broad features of $f(a)$, in the form of a non-negative function $p(a)$, such that $p(a) = 0$ only if $f(a) = 0$; of course, in practice we will not know where the support of f is, exactly, so $p(a)$ should be positive everywhere. A rough idea of the support of f can be indicated by concentrating p in that region. Let p play the role of R_{xx}, f the role of R_{xz}; we are effectively assuming that, for some $\varepsilon > 0$, we have $p(a) \geq \varepsilon f(a)$ for all a, and that (apart from the scaling) $p(a)$

overestimates $f(a)$ everywhere; the scale factor cancels in the end, so is not needed. Then $H = f/p$ is approximated by the polynomial D^M_{-M} and the equations that must be solved are

$$F(m) = \sum_{k=-M}^{M} d(k)\, P(m-k) \qquad -M \leqslant m \leqslant M \tag{3.1}$$

where $P(m)$ are the Fourier coefficients of $p(a)$. These equations arise when we minimise the approximation error

$$\int_{-\pi}^{\pi} \left| f(a) - p(a) \sum_{k=-M}^{M} d(k) \exp(ika) \right|^2 p(a)^{-1}\, da \tag{3.2}$$

as a function of the $d(k)$. The resulting estimator of $f(a)$ is the PDFT [6], so called because of its form:

$$\text{PDFT}(a) = p(a) D^M_{-M}(a). \tag{3.3}$$

If the prior estimate $p(a) =$ constant, $|a| \leqslant \pi$, then $d(k) = F(k)$ and the PDFT reduces to the DFT.

If the data are oversampled relative to the actual support of $f(a)$ then including information about this support in the $p(a)$ can result in significant improvement, so long as regularisation to avoid sensitivity to noise is used [1]. Note that, although the PDFT is not necessarily non-negative, it is data consistent; it extrapolates values of $F(m)$ beyond the data window.

3.2. *Nonlinear methods*

We assume now that $f(a)$ consists of two components, a discrete (delta functions) component, which is the object of interest, and a background (continuous) component, about which we have some prior information; let $p(a)$ be our non-negative prior estimate of the background component. Letting $p(a)$ play the role of R_{ss} and $f(a)$ the role of R_{xx}, we see that $H = p/f$; for the filter function D^L_K to remove from $R_{xx} = f$ the component associated with $R_{uu} = f - p$ it must place nulls near the values of the support of the discrete component. We perform the calculations to obtain the D^L_K and then examine the nulls. We have some freedom in the choice of the K and L; two choices are of particular importance: (i) $M/2 \geqslant L = -K$; (ii) $L = M$, $K = 0$. The first has as a special case the symmetric linear predictor (SLP) [7, 8], while the second includes Burg's MEM.

In case (i) we solve the equations

$$P(m) = \sum_{k=-L}^{L} d(k)\, F(m-k) \qquad |m| \leqslant L = M/2 \quad (M \text{ even}) \tag{3.4}$$

and use the fact that D^L_{-L} approximates $H = p/f$ to obtain, as the estimate of f, the centred inverted PDFT (CIPDFT):

$$\text{CIPDFT}(a) = p(a)/D^L_{-L}(a). \tag{3.5}$$

If the prior $p(a) =$ constant, then (3.5) becomes the symmetric linear predictive method of Johnson [7]. Note that if the support of f is properly contained within the support of p then, in order to obtain the $d(k)$ that are optimal for that f and p it is

necessary to replace $P(m)$ in (3.4) by the corresponding Fourier coefficient of the function that is $p(a)$ on the support of f and zero otherwise. In practice one does not know the support of f; our point is rather that (3.4) does not provide the optimal $d(k)$ for such pairs p and f.

The role of the prior $p(a)$, in the nonlinear methods, is to reduce bias in the estimate of peak locations; if we estimate the background component badly then the filter, as it tries to eliminate the $f(a)-p(a)$ features, must null out (true background $-p$) as well as the discrete component. With limited freedom to place nulls, bias is unavoidable. This has been shown to be a problem with MEM, when used on oversampled data [1], and is due to the assumption, implicit in MEM, that the background is constant over $[-\pi, \pi]$. We consider MEM next, as a special case of (ii).

For case (ii) we have $K=0$, $L=M$ and we solve equations

$$P(m)=\sum_{k=0}^{M} d(k)\,F(m-l) \qquad 0\leq m\leq M \tag{3.6}$$

to obtain the filter function D_0^M; we view this function now as an approximation of G, not of $H=p/f$. The discrete Wiener–Hopf equations (2.4) are equivalent to the statement $(R_w)_+=(R_{ff}G)_+$, where by $(R_w)_+$ we mean the causal part of the Fourier series

$$(R_w)_+(a)=\sum_{m=0}^{\infty} r_w(m)\exp(ima) \qquad |a|\leq\pi \tag{3.7}$$

and similarly for other functions. Equations (3.6) tell us that the two causal functions p_+ and $(fD_0^M)_+$ have identical Fourier coefficients, out to index $m=M$. Because D_0^M is a finite polynomial we can rewrite $fD_0^M)_+$ as $(fD_0^M)_+=f_+D_0^M+j_+$, where j_+ is a finite causal polynomial involving only known values:

$$j_+(a)=\sum_{m=1}^{M-1}\left(\sum_{k=1}^{M-m} F(-k)\,d(m-k)\right)\exp(ima) \tag{3.8}$$

From $p_+=(fD_0^M)_+=f_+D_0^M+j_+$ we obtain an estimate q of f_+;

$$q(a)=(p_+(a)-j_+(a))/D_0^M(a); \tag{3.9}$$

from $f=2\mathrm{Re}(f_+)-F(0)$ we obtain the inverse PDFT (IPDFT) estimate of f itself:

$$\mathrm{IPDFT}(a)=2\mathrm{Re}(q(a))-F(0). \tag{3.10}$$

Consider the complex polynomial $D(z)=d(0)+d(1)z+\ldots+d(M)z^M$. If the roots of $D(z)$ are outside the unit circle (the minimum phase, or MP property) then $1/D_0^M(a)$ is also causal and so is $q(a)$. It can be shown easily that, if $1/D_0^M(a)$ is causal, then IPDFT(a) is data consistent, although it may not be non-negative. Although it is not always the case that $D(z)$ has the MP property, it is frequently the case in practice, and the IPDFT is usually data consistent. If the prior $p(a)=$ constant, then the IPDFT reduces to Burg's MEM, the $D(z)$ has the MP and the MEM is data consistent (as well as non-negative).

As remarked earlier, the MEM has been observed to perform poorly when the function f is concentrated in a smaller region of $[-\pi, \pi]$; this is because the $p(a)$ is a constant, while the background is not evenly distributed over all of $[-\pi, \pi]$. Because the IPDFT is free to take on negative values it could be used to gauge the accuracy of

the prior being used; significant negative values should indicate, that our $p(a)$ is not accurate. We have not obtained a quantitative measure of the significance of negative values, however.

We consider now the extensions of these methods to the two-dimensional case. Although much remains essentially the same the absence of an obvious generalisation of the notion of causality affects the extension of the IPDFT.

4. The two-dimensional case

As in the one-dimensional case, the power spectrum of the input, $R_{xx}(\alpha, \beta)$, is the sum of two components, $R_{xx}(\alpha, \beta) = R_{ss}(\alpha, \beta) + R_{mm}(\alpha, \beta)$, and the Wiener filter is the doubly indexed sequence $\{h(j, k)\}$ of Fourier coefficients of the function $H(\alpha, \beta) = R_{ss}(\alpha, \beta)/R_{xx}(\alpha, \beta)$. To obtain a finite-impulse-response approximation to the Weiner filter we minimise the following error of approximation, as a function of the $d(j, k)$:

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |H(\alpha, \beta) - D_{I,K}^{J,L}(\alpha, \beta)|^2 R_{xx}(\alpha, \beta)\, d\alpha d\beta \tag{4.1}$$

where

$$D_{I,K}^{J,L}(\alpha, \beta) = \sum_{j=I}^{J}\sum_{k=K}^{L} d(j, k) \exp(ij\alpha + ik\beta). \tag{4.2}$$

In the two-dimensional case the problem is to reconstruct the non-negative function $f(a, b)$, $|a|, |b| \leq \pi$, from finitely many values of its Fourier coefficients, $F(m, n)$, $|m| \leq M$, $|n| \leq N$. As in the one-dimensional case we consider estimates of $f(a, b)$ obtained by analogy with the problem of approximating the Wiener filter.

4.1. Linear methods

Assume that a prior estimate of the general shape of $f(a, b)$ is given by the positive function $p(a, b)$, and that $\{P(m, n)\}$ are its Fourier coefficients. As before, we let p play the role of R_{xx}, f the role of R_{ss}, so that the Wiener filter is $H = f/p$.

For fixed I, J, K, L the optimal finite-impulse-response filter function is $D(\alpha, \beta) = D_{I,K}^{J,L}(\alpha, \beta)$ where the coefficients of D satisfy the equations

$$F(m, n) = \sum_{j=I}^{J}\sum_{k=K}^{L} d(j, k)\, P(m-j, n-k) \qquad |m| \leq M, \quad |n| \leq N. \tag{4.3}$$

Having found the $d(j, k)$ we use the fact that D approximates $H = f/p$ to obtain our estimate of f:

$$\text{PDFT}(a, b) = p(a, b)\, D(a, b). \tag{4.4}$$

The obvious choices for I, J, K, L are $-I = J = M$, $-K = L = N$.

4.2. Nonlinear methods

We assume now that $f(a, b)$ has a discrete component of interest, as well as a background component estimated by the non-negative function $p(a, b)$. As in the 1D

case we let p play the role of R_{ss}, f the role of R_{ss}, so that the approximate Wiener filter attempts to null out the discrete component. If the support of f contains the support of p then the equations to be solved for the optimal finite-impulse-response filter $\{d(j, k)\}$ are

$$P(m, n) = \sum_{j=I}^{J} \sum_{k=K}^{L} d(j, k)\, F(m-j, n-j) \qquad I \leq m \leq J, \quad K \leq n \leq L. \tag{4.5}$$

The choices for I, J, K, L will be restricted by the available data, since the data make up the entries of the matrix that appears in the system of equations to be solved. We consider here two possibilities.

(i) Let $J = -I = M/2$, $L = -K = N/2$ (M, N even). Solving (4.4) for the $d(j, k)$ we view $D = D^{J,L}_{I,K}$ as an approximation of $H = p/f$, so that our estimate of f is the two-dimensional version of (3.5):

$$\text{CIPDFT}\,(a, b) = p(a, b)/D(a, b). \tag{4.6}$$

(ii) Let $I = K = 0$, $J = M$, $L = N$. Then D can be viewed as an estimate of the first-quadrant-indexed component of H, which we denote by H_{++}. With the first-quadrant-indexed component of $p(a, b)$ given by

$$p(a, b)_{++} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P(m, n) \exp(ima + inb) \tag{4.7}$$

equations (4.4) state that $p(a, b)_{++}$ and $[f(a, b)D(a, b)]_{++}$ have the same Fourier coefficients, for indices $0 \leq m \leq M$, $0 \leq n \leq N$. As in the 1D case, we can write $[f(a, b)D(a, b)]_{++} = f(a, b)_{++}D(a, b) + j(a, b)_{++}$, where $j(a, b)_{++}$ is a first-quadrant-indexed function that involves only known values. Our estimate of $f(a, b)_{++}$ is then $q(a, b) = [p(a, b)_{++} - j(a, b)_{++}]/D(a, b)$, which may not itself be first-quadrant-indexed. Repeating this procedure three more times, for each quadrant, we estimate $f(a, b)$ by summing the four estimates so obtained, taking care to subtract components included in more than one estimator. The resulting estimator we call the IPDFT.

5. Relation to other methods

The reconstruction problem considered here is to obtain the function $f(a, b)$ from the values $F(m, n)$, $|m| \leq M$, $|n| N$, where

$$F(m, n) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(a, b) \exp[-(ima + inb)]\, da\, db/4\pi^2; \tag{5.1}$$

that is, we are attempting to solve the integral equation. The survey paper by Frieden [9] describes a number of approaches to this problem.

When the $p(a, b)$ is chosen to incorporate support information, so that $p(a, b) = 1$, $|a| < A < \pi$, $|b| < B < \pi$, and $p(a, b) = 0$ elsewhere, a small amount of noise in the data can cause degradation of the PDFT estimator. It is safer to make $p(a, b) = \varepsilon > 0$, instead of $p(a, b) = 0$. This is a form of regularisation and is in keeping with the requirement that the support of p be no smaller than that of f. Since the PDFT performs an approximation of the function f in the (a, b) domain and smooths the effects of noise (if regularised), it resembles the methods of Phillips [10] and Twomey [11]. The main

differences are that the PDFT retains the continuous formulation, rather than discretising $f(a, b)$, and employs a prior estimate of the function f.

It might appear that the PDFT is related to the Helstrom–Weiner 'sharpness-constrained' method [12]. The latter is based, however, on a statistical, or ensemble, model for the restoration problem and employs power spectra of f and p; the mean squared error is calculated in the usual L^2 norm, rather than with a weighted norm in (a, b) space. In the approach presented here we do not postulate the existence of an ensemble of object functions f to be restored and the idea of Wiener filtering is introduced only to borrow the weighted error criterion used for approximating non-negative functions by polynomials. The Backus–Gilbert [13] method is similar in philosophy to the Helstrom–Weiner approach but there is only a superficial connection to the estimators presented here.

Because the finite data are typically insufficient to determine a single, unique solution to the reconstruction problem, one is faced with the task of selecting, from among the many possibilities, one particular answer. The general feeling, which we share, is that the selection should not be arbitrary but should be guided by some reasonable principles of inference. At this point there is some disagreement concerning which principles of inference are to be called reasonable. In an attempt to resolve the situation Shore and Johnson [14] developed an axiomatic basis for the principle of cross-entropy minimisation and Jones [15] has recently provided an approximation-theoretic argument for the same method.

Among all functions $g(a, b)>0$ consistent with our data we could select the one for which the Shannon entropy

$$\text{entropy}(g) = -\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(a, b)\log g(a, b)\,da\,db \tag{5.2}$$

is maximised. Generally, there is no closed-form solution and iterative procedures are employed.

If there is available a prior estimate, $p(a, b)$, of $f(a, b)$ then (5.2) is replaced by the cross-entropy of g, given p:

$$\text{cross-entropy}(g|p) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(a, b)\log[g(a, b)/p(a, b)]\,da\,db \tag{5.3}$$

The method of 'minimisation of cross-entropy' (MCE) has us select, as the estimate of $f(a, b)$, that data-consistent $g(a, b)>0$ for which the integral in (5.3) is minimised. The optimal solution then has the form

$$\text{MCE}\,(a, b) = p(a, b)\exp\left(\sum_{j=-M}^{M}\sum_{k=-N}^{N} c(j, k)\exp(ija + ibk)\right) \tag{5.4}$$

If $p(a, b)$ is a good prior estimate then the sum in the exponential term will be near zero; approximating $\exp(x)$ by $1+x$ leads to an estimator of the PDFT form. If it is known that the function $f(a, b)$ is spiky, then the sum in the exponential term will have significant negative values. If we estimate $\exp(x)$ by $1/(1-x)$, then this is better than $1+x$ for negative x; making this approximation in (5.4) leads to the CIPDFT form.

When the $p(a, b)$ is constant over the support of the object function $f(a, b)$ the PDFT (4.4) provides a minimum energy extrapolation of the data, consistent with the support constraint. In reference [16] we considered the problem of reconstructing $f(a,$

b) from only the magnitude data, $|f(m, n)|$, $|m| \leq M$, $|n| \leq N$, that is, the phase retrieval problem. When arbitrary phases are assigned to the magnitude data and the PDFT energy calculated one finds the energy to be dependent on that choice of phases and therefore to provide a useful cost function to direct the search for the correct phases.

6. Conclusions

In this paper we have considered the reconstruction of a non-negative function from finitely many values of its Fourier transform. We have extended to the 2D case methods previously presented for 1D reconstruction [1] and obtained a new derivation of these estimators based on analogy with the design of approximate Wiener filters, in which the object function to be reconstructed and our prior estimate play the roles of input and output power spectra. To obtain linear estimators we let our prior estimate play the role of the input power spectrum, allowing the filter to extract those features not found in the true object. To obtain nonlinear estimators for spiky objects with continuous backgrounds we estimate the background function, and then let it play the role of the output power spectrum; the true object function then plays the role of the input power spectrum, so that the filter attempts to null out the discrete component.

The linear methods are extrapolation procedures that are particularly useful when the data are oversampled. The nonlinear methods generalise the Burg maximum entropy method, for the 1D case, and provide computationally inexpensive 2D approximations to other entropy-based methods.

The methods presented here are derived using the best approximation in weighted Hilbert spaces; the linear equations to be solved in each case are the *normal equations* that arise from such a best approximation and we know what is being approximated by what in each case.

Acknowledgments

The authors acknowledge the financial support of the Office of Naval Research, through Contract N00014-85-K-0402 and NATO grant 725/84. They also wish to thank Professor James Howard and the staff of the Human Performance Laboratory, The Catholic University of America, for their assistance.

References

[1] Byrne C L, Fitzgerald R M, Fiddy M A, Hall T J and Darling A M 1983 Image restoration and resolution enhancement *J. Opt. Soc. Am.* 73 1481–7

[2] Ekstrom M 1982 Realisable Wiener filtering in two dimensions *IEEE Trans. Acoust., Speech Signal Proc.* **As-30** 31

[3] Papoulis A 1977 *Signal Analysis* (New York: McGraw-Hill)

[4] Burg J P 1975 Maximum entropy spectral analysis *PhD Dissertation* Stanford University

[5] Byrne C L and Fitzgerald R M 1983 An approximation theoretic approach to maximum entropy spectral analysis *IEEE Trans. Acoust., Speech Signal Proc.* **As-31** 734–6

[6] Byrne C L and Fitzgerald R M 1984 Spectral estimators that extend the maximum entropy and maximum likelihood methods *SIAM J. Appl. Math.* **44** 425–42

[7] Johnson D 1982 The application of spectral estimation methods to bearing estimation problems *Proc. IEEE* **70** 1018–28

[8] Cox H, Zeskind R and Owen M 1985 The effect of random errors on linear predictive array processors *Bolt Beranek and Newman, Inc. Technical Report No* 5595

[9] Frieden B R 1978 Image enhancement and restoration *Picture Processing and Digital Filtering* (*Topics in Applied Physics* vol. 6) ed. T S Huang (New York: Springer)

[10] Phillips D L 1962 A technique for the numerical solution of certain integral equations of the first kind *J. Assoc. Comput. Mach.* **9** 84–97

[11] Twomey S 1963 On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature *J. Assoc. Comput. Mach.* **10** 97–101

[12] Helstrom C W 1967 Image restoration by the method of least squares *J. Opt. Soc. Am.* **57** 297–303

[13] Backus G and Gilbert F 1970 Uniqueness in the inversion of gross earth data *Phil. Trans. R. Soc.* A **266** 123–92

[14] Shore J E and Johnson R W 1980 Axiomatic derivation of the principle of maximum entropy and minimum cross entropy *IEEE Trans. Inform. Technol.* **IT–26** 926–43

[15] Jones L K 1987 Approximation theoretic uniqueness for the principle of minimum cross entropy *IEEE Trans. Inform. Technol.* to be published

[16] Byrne C L and Fiddy M A 1987 Estimation of continuous object distributions from limited Fourier magnitude measurements *J. Opt. Soc. Am.* A **4** 112–7

draft

Stable data adaptive matched field methods for ambiguity reduction in source parameter estimation

by

Charles L. Byrne and Christopher Feuillade

(First draft)

ABSTRACT

Data adaptive methods, such as Capon's maximum likelihood (ML) method, suppress sidelobe structure and reduce ambiguity in source parameter estimation because they are optimized against unwanted terms actually present in the data, rather than against an a priori model of what could be present. When the noise component resembles potential source terms and produces a reduced rank cross-sensor correlation matrix in the noise-only case methods such as ML can become unstable. By employing a "reduced rank" ML estimator we can avoid this instability. This method is analogous to the "sector-focused stability" method recently developed by Byrne and Steele, and is derived by considering the general problem of suppressing ambiguity in parameter estimation.

The generalized sidelobe problem and optimal suppression

A problem that arises often in applications is the following one, stated here in general terms:

The matching problem: Let Θ be a family of (possibly vector) parameters θ, and let $\mathcal{P}=\{\underline{p}(\theta) \mid \theta \text{ in } \Theta\}$ be a family of pairwise linearly independent N-dimensional vectors parametrized by the θ in Θ. Our data consists of the single vector $\underline{p}=\underline{p}(\theta_0)$, taken from $\mathcal{P}$ and the problem is to determine the value $\theta=\theta_0$.

If the members of $\mathcal{P}$ are quite distinct and Θ is a small finite set then the problem is easily solved by inspection. More commonly, the set Θ is a continuum, the vector function $\underline{p}(\theta)$ continuous in θ, and the data vector a noisy version of $\underline{p}(\theta_0)$. The usual approach in such cases is to perform linear filtering, such as simple matching via the dot product, and base the decision on the outcomes of the filtering.

Linear filtering solution: Consider each member θ of Θ in turn, hold θ fixed and select a linear filter $\underline{f}=\underline{f}(\theta)$, a N-dimensional vector with entries dependent on θ. Let $y(\theta)=\left|\underline{f}^*\underline{p}\right|^2=\underline{f}^*\underline{p}\underline{p}^*\underline{f}$ be the (magnitude squared of the) filter output. We know from Cauchy's inequality that $y(\theta)\le(\underline{p}^*\underline{p})(\underline{f}^*\underline{f})$, with equality if and only if $\underline{f}=\alpha\underline{p}$, for some scalar α: from the function $y(\theta)$ we can then determine θ_0.

Example: For each θ let $\underline{f}(\theta)=\underline{p}(\theta)\,/\,\sqrt{(\underline{p}(\theta)^*\underline{p}(\theta))}$. Then $y(\theta)\le\underline{p}^*\underline{p}$, with equality if and only if $\theta=\theta_0$. This method we shall call simple matching: the graph of the function $y(\theta)$ is usually called the ambiguity surface.

If, in the simple matching approach, the output $y(\theta_1)$ is large for a value of $\theta_1\neq\theta_0$, we say that there is a sizable sidelobe at θ_1, for the true value θ_0. We shall describe two general procedures for reducing the sidelobe effect, the first a linear method that does not depend on the actual data obtained, and the second a data adaptive method analogous to Capon's maximum likelihood method (MLM) for spectrum estimation.

A linear filtering method to optimally reduce sidelobes

Denote by $y(\theta;\theta_0)$ the magnitude squared of the filter output $\underline{f}(\theta)^*\underline{p}$ corresponding to fixed value θ and true value θ_0. For any particular problem we do not know the value of θ_0, which could be any member of Θ. For those $\theta \neq \theta_0$ we want $y(\theta;\theta_0)$ to be small; since we do not know θ_0 let us make the average value of $y(\theta;\theta_0)$ small, as θ_0 ranges over the various members of Θ. Specifically, for fixed θ, select that filter $\underline{f}(\theta)$ for which $\int y(\theta;\theta_0)/\underline{p}(\theta_0)^*\underline{p}(\theta_0)\, d\theta_0$ is minimized, subject to the constraint $y(\theta;\theta)=1$; the precise meaning of the integral will be clear from the context. We can formulate this in matrix language as follows:

Minimize $\underline{f}^*A\underline{f}$, subject to $\underline{f}^*\underline{p}(\theta)=1$, where A is the N by N matrix with entries $A_{m,n}= \int \underline{p}(\theta_0)_m\, \underline{p}(\theta_0)^*_n /\underline{p}(\theta_0)^*\underline{p}(\theta_0)\, d\theta_0$, and subscripts denote the particular entry of the matrix or vector. Using the normalized vectors $\underline{u}(\theta_0)= \underline{p}(\theta_0)/\sqrt{(\underline{p}(\theta_0)^*\underline{p}(\theta_0))}$ we can write $A= \int \underline{u}(\theta_0)\underline{u}(\theta_0)^*\, d\theta_0$. The optimal filter is then $\underline{f}_{opt}(\theta)= \lambda(\theta)A^{-1}\underline{p}(\theta)$, where the parameter $\lambda(\theta)= 1/\underline{p}(\theta)^*A^{-1}\underline{p}(\theta)$, provided that A is invertible. The value $OPT(\theta)=|\underline{f}_{opt}(\theta)^*\underline{p}|^2 = |\underline{p}(\theta)^*A^{-1}\underline{p}|^2/|\underline{p}(\theta)^*A^{-1}\underline{p}(\theta)|^2$ is then the function we want to use to compute the optimized ambiguity surface. If the matrix A is not invertible then pseudo-inversion is used to obtain the optimal filter; this will be the case in the normal-mode situation considered below.

For fixed θ the optimal filter $\underline{f}(\theta)$ operates on the data vector $\underline{p}$ and we want the value $|\underline{f}(\theta)^*\underline{p}|^2$ to be small if $\theta \neq \theta_0$. The optimal filter is designed to make this value small, on average, but the actual data we have is a particular $\underline{p}(\theta_0)$; we do not really care if $|\underline{f}(\theta)^*\underline{p}(\theta_1)|^2$ is large for other values $\theta_1 \neq \theta_0$, since $\underline{f}(\theta)$ does not have to operate on $\underline{p}(\theta_1)$. Data adaptive methods, such as Capon's MLM, optimize the filter against what is actually there in the data, not against a class of potential, but mostly not actual, data vectors. We consider next the extension to our general problem of the data adaptive approach of Capon.

Nonlinear data adaptive filtering for sidelobe suppression

In practice the data may consist of several $\underline{p}$, each a noisy version of $\underline{p}(\theta_0)$. Of interest then is the average over all the $\underline{p}$ of the value of the output $y(\theta:\theta_0)$: average output = $<|\underline{f}(\theta)^*\underline{p}|^2> = \underline{f}(\theta)^*<\underline{p}\underline{p}^*>\underline{f}(\theta) = \underline{f}(\theta)^*R\underline{f}(\theta)$, where $<\ >$ denotes averaging over the available $\underline{p}$ and R is the matrix $R=<\underline{p}\underline{p}^*>$. For fixed θ let us find that filter $\underline{f}(\theta)$ for which this averaged output, $\underline{f}(\theta)^*R\underline{f}(\theta)$, is minimized, subject to the constraint $\underline{f}(\theta)^*\underline{p}(\theta)=1$. This "maximum likelihood" solution is easily seen to be $\underline{f}_{ml}(\theta)= \lambda(\theta)R^{-1}\underline{p}(\theta)$, where the parameter is $\lambda(\theta) = 1/\underline{p}(\theta)^*R^{-1}\underline{p}(\theta)$. The averaged output is then $ML(\theta)= 1/\underline{p}(\theta)^*R^{-1}\underline{p}(\theta)$; here it is assumed that one effect of averaging and of noise is to make the rank of R equal to N. Note that $OPT(\theta)$ differs from $ML(\theta)$ in that A is replaced by R in ML: it is in this sense that ML is data adaptive.

The ML approach will generally outperform the OPT method because it can employ its algebraic freedom to reduce sidelobe effects coming from what is actually present in the data, rather than to guard against potential threats that are most likely not present.

Because loss of resolution in linear estimation is a particular form of sidelobe problem the ML approach typically achieves better resolution than linear methods. If the data vector is a superposition of two or more members of $\mathcal{P}$, say $\underline{p}=\underline{p}(\theta_0) + \underline{p}(\theta_1)$, with θ_0 and θ_1 similar, then the simple matching method may result in $|\underline{f}(\theta)^*\underline{p}|^2$ being largest at $\theta=\theta_2$, where θ_2 is neither θ_0 nor θ_1, but is near each. We could say that in this case the sidelobe at θ_2 caused by θ_1 is added to that caused by θ_0, resulting in a maximum between the two correct values. Because ML can do a better job of suppressing these sidelobes it can resolve when the simple matching, or even the OPT method, cannot.

In many practical situations the data vector includes an additive random noise component. Depending on the statistical behavior of this noise component the performance of the ML method can vary considerably. In [ ] we discussed the instability of the ML estimator of bearing of planewave source fields, in the presence of spatially concentrated noise and systematic phase errors. The instability arises when the noise component "looks like" potential signals and corresponds to a reduced rank component of the matrix R.

In the next section we shall consider the OPT and ML methods for the case of normal mode propagation in shallow water; here also the noise component can lead to instability of the ML method and we shall need to develop more stable procedures.

Normal-mode propagation and signal processing

The field at range r=0 and depth z excited by a unit amplitude point source at range $r=r_0$ and depth z_0 in a waveguide is

$$P(z) = P(z;r_0,z_0) = \pi i \sum_{m=1}^{M} S_m(r_0,z_0)\, U_m(z), \qquad (1)$$

where $U_m(z)$ is the mth modal eigenfunction of the depth-dependent boundary value problem, and $S_m(r_0,z_0)$ is the modal amplitude value, given by

$$S_m(r_0,z_0) = \exp(3\pi i/4)\exp(-\beta_m r_0 + ik_m r_0) U_m(z_0) \sqrt{(2\pi/k_m r_0)} \; . \qquad (2)$$

The pressure field is sampled using a vertical array of sensors, at depths z_n, n=1,...,N, and single frequency components extracted via FFT to produce the vector $\underline{p}=(P(z_1),....,P(z_N))^T$. This vector, obtained (ideally) from sampling the signal-only field, can be written as $\underline{p}=U\underline{s}$, where U is the N by M matrix with entries $U_{n,m}=U_m(z_n)$, and $\underline{s}$ the M by 1 vector with entries $\underline{s}_m=S_m(r_0,z_0)$.

In the case of a high-loss bottom most of the noise reaching the array is unaffected by the bottom, so contributes to $\underline{p}$ a vector of the form $\underline{n}$, whose entries are (possibly correlated) random variables. If the bottom is low-loss then noise energy can excite the modal structure and contribute to $\underline{p}$ a component of the form $U\underline{\gamma}$, where $\underline{\gamma}$ is an M by 1 vector whose entries are random variables representing the aggregate excitation of each mode by the superposition of noise sources. We can therefore write the data vector as $\underline{p} = U\underline{s} + U\underline{\gamma} + \underline{n} = U\underline{x} + \underline{n}$, with randomness entering through the $\underline{\gamma}$ and the $\underline{n}$; when one considers the effects of rough surface scattering on the signal one includes random phase and amplitude modulations of the entries of $\underline{s}$. The matrix

$R = \langle \underline{p}\underline{p}^* \rangle$ now has the form

$$R = U\langle \underline{s}\underline{s}^* \rangle U^* + U\langle \underline{\gamma}\underline{\gamma}^* \rangle U^* + \langle \underline{n}\underline{n}^* \rangle = U\underline{s}\underline{s}^* U^* + UQU^* + G \quad (3)$$

We shall assume that the non-modal noise component, G, is a multiple of the identity matrix, representing spatially uncorrelated noise, such as sensor noise. The interesting term in (3) is UQU^*, the modal noise component of R.

Typically the number of modes, M, will be less than the number of sensors, N, so that the rank of the matrix UQU^* is at most M. Because the signal component, $U\underline{s}$, and the noise component, $U\underline{\gamma}$, both lie in the M-dimensional linear span of the M columns of U the noise looks like potential signals and we expect ML to exhibit the sort of instability we discussed earlier. We shall return to this point when we discuss stable nonlinear methods; for the moment we consider the OPT method in the context of normal modes.

Optimal linear processing in the normal mode case: Let us denote by $\underline{p}(\theta_0) = \underline{p}(r_0, z_0)$ the vector of field samples we would obtain in the case of only a single source at $\theta_0 = (r_0, z_0)$. Let us normalize $\underline{p}(\theta_0)$ to get $\underline{u}(\theta_0) = \underline{p}(\theta_0)/\sqrt{(\underline{p}(\theta_0)^* \underline{p}(\theta_0))}$. The matrix A that represents the totality of potential source vectors is now $A = \int \underline{u}(\theta_0)\underline{u}(\theta_0)^* \, d\theta_0$, where integration with respect to θ_0 means over $0 \le z_0 \le H$=channel depth, and over $0 \le r_0 < \infty$, and $d\theta_0 = r_0 dr_0 dz_0$. From (1) and (2) it follows that the matrix A has the form $A = UBU^*$, where B is the matrix with entries $B_{m,k}$ given by

$$B_{m,k} = \int S_m(r_0, z_0) S_k(r_0, z_0)^* \, \| \underline{p}(r_0, z_0) \|^{-2} \, r_0 dr_0 dz_0 \, . \quad (4)$$

and $\| \underline{p}(r_0, z_0) \| = \sqrt{(\underline{p}(r_0, z_0)^* \underline{p}(r_0, z_0))}$.

The minimization problem to be solved is: Minimize $\underline{f}^* A \underline{f}$, subject to $\underline{f}^* \underline{p}(\theta) = 1$. This is equivalent to: Minimize $\underline{v}^* B \underline{v}$, subject to $\underline{v}^* \underline{s}(r, z) = 1$.

where $\underline{v}=U^*\underline{f}$ and $\underline{p}(r,z)=U\underline{s}(r,z)$. The optimal $\underline{v}$ is $\underline{v}_{opt}=\alpha B^{-1}\underline{s}(r,z)$ for $\alpha=1/\underline{s}(r,z)^*B^{-1}\underline{s}(r,z)$. The optimal filter $\underline{f}_{opt}$ is found as follows: assume that $\underline{f}_{opt}=U\underline{w}$ for some $\underline{w}$; then $\underline{v}_{opt}=U^*\underline{f}_{opt}=U^*U\underline{w}$, so that $\underline{w}=(U^*U)^{-1}\underline{v}_{opt}$ and $\underline{f}_{opt}=U(U^*U)^{-1}\underline{v}_{opt}$. The optimal estimator is then

$$OPT(r,z)=<\left|\underline{f}_{opt}{}^*\underline{p}(r,z)\right|^2>=\alpha^2\underline{s}(r,z)^*C^*RC\underline{s}(r,z), \qquad (5)$$

where $C=U(U^*U)^{-1}B^{-1}$.

Note that the entries of U^*U are $(U^*U)_{m,j}=\sum_{n=1}^{N} U_m(z_n)U_j(z_n)^*$.

Even if $\int_{z=0}^{H} U_m(z)U_j(z)^*dz=0$ for $m\neq j$, it may not happen that $(U^*U)_{m,j}=0$ for $m\neq j$; the latter involves the locations of the sensors, whereas the former does not. Because the integral is only over the water column and does not include the bottom it may not be zero either, as in cases such as the Pekeris model in which $0\leq z<\infty$ are the limits on z. (6)?

Mode-filtering as an approximation to the optimal processor: We obtain the mode filtering procedure discussed in Shang[ ] by making several simplifying assumptions in the computation of the matrix B: specifically, let us assume that $||\underline{p}(r_0,z_0)||$ is constant, as a function of (r_0,z_0); and that the functions $U_m(z_0)$ are orthogonal over the interval $0\leq z_0<H$. Then the entries of B are $B_{m,j}=0$, $m\neq j$, $B_{m,m}=1/2\beta_m k_m$, so that B is a diagonal matrix. If we assume, in addition, that the values β_m and k_m do not vary greatly with m, then B is (a multiple of) the identity matrix and so $C=U(U^*U)^{-1}$. The OPT estimator is then approximated by the mode filtering estimator (MFE)

$$MFE(r,z)=\alpha'^2\underline{s}(r,z)^*(U^*U)^{-1}U^*RU(U^*U)^{-1}\underline{s}(r,z)\,. \qquad (7)$$

with $\alpha'^2=\underline{s}(r,z)^*\underline{s}(r,z)$.

The maximum likelihood approach in the normal-mode case: From the matrix $R=\langle \underline{p}\underline{p}^* \rangle$, we obtain the ML estimate:

$$ML(r,z)= 1/\ \underline{p}(r,z)^* R^{-1} \underline{p}(r,z)\ . \qquad (8)$$

Since $R=U\underline{s}\underline{s}^*U^* + UQU^* + \varepsilon I$, it follows that, if the rank of Q is M, the eigenvalues λ_n of R, for n=M+1,...,N, satisfy $\lambda_n=\varepsilon$ and that the associated eigenvectors $\underline{x}_n$ have the property that $U^*\underline{x}_n=\underline{0}$: for n=1,...,M $\lambda_n>\varepsilon$. Writing $R = XLX^*$, where the columns of X are the orthonormal eigenvectors $\underline{x}_n$, and L is the diagonal matrix of eigenvalues of R, that is $L =diag(\lambda_1,...,\lambda_N)$, we can rewrite (8) as

$$ML(r,z)= 1/\ \sum_{n=1}^{N} \lambda_n^{-1}\ \left|\underline{p}(r,z)^*\underline{x}_n\right|^2 \ = 1/ \sum_{n=1}^{N} \lambda_n^{-1} \left|\underline{s}(r,z)^*U^*\underline{x}_n\right|^2 . \qquad (9)$$

Because of the reciprocal weighting by λ_n^{-1} those terms corresponding to the lowest eigenvalues contribute most; that is, the sum is essentially over n=M+1,...,N. Because, for these n, $\underline{p}(r_0,z_0)^*\underline{x}_n =0$ for (r_0,z_0) associated with the true source, we might expect, as in the usual case of ML estimation, to discover the value of (r_0,z_0) by evaluating ML(r,z) and looking for the largest value. However, each of the terms in (9) corresponding to n=M+1,...,N is zero, for all (r,z). It follows that the ML estimator will show a large response for each (r,z), not only for the (r_0,z_0) corresponding to the actual source. The resulting ambiguity surface will be essentially uniformly "white".

It is important to recall that this failure of the ML estimator occurs when the noise component is essentially modal, that is, of the form UQU^*, as would be the case for a low-loss bottom; if there is no noise component of this form then the ML procedure should work better. When the modal noise is present the situation is analogous to that of spatially concentrated planewave noise and the solution we shall consider next is similar to the "sector-focused stability" (SFS) method given in [ ].

Reduced-dimension ML method for the normal-mode case

The matrix U is N by M and induces a linear transformation from the space of complex M-dimensional vectors, C^M, into the space of complex N-dimensional vectors, C^N; we shall assume that U^*U is invertible, which is equivalent to U being one-to-one, or to U^* being onto C^M. Every $\underline{x}$ in C^N can be written uniquely as a sum $\underline{x}=U\underline{v}+\underline{w}$, for some $\underline{v}$ in C^M and $\underline{w}$ in C^N such that $U^*\underline{w}=\underline{0}$; in fact, $\underline{v}=(U^*U)^{-1}U^*\underline{x}$ and $\underline{w}=\underline{x}-U\underline{v}$.

The ML method, because it relied on the eigenvectors of R associated with the lowest eigenvalues, failed when those eigenvectors $\underline{x}_n$, n=M+1,...,N, had the property $U^*\underline{x}_n=\underline{0}$. To obtain a data adaptive, nonlinear method that works in the presence of modal noise we need to rely on vectors $\underline{x}$ with the property that $\underline{p}(r_0,z_0)^*\underline{x}=0$, but such that $\underline{p}(r,z)^*\underline{x}\neq 0$ for other values of (r,z); in particular, we must not have $U^*\underline{x}=\underline{0}$, or even near zero. One way to prevent this behavior is to require that $\underline{x}$ not have the component $\underline{w}$ such that $U^*\underline{w}=\underline{0}$; that is, require that $\underline{x}$ have the form $\underline{x}=U\underline{v}$ for some $\underline{v}$. Let us now solve the following optimization problem, in lieu of the one that leads to the lowest eigenvalue of R:

Minimize $\underline{x}^*R\underline{x}$, subject to $\underline{x}\underline{x}^*=1$ *and* $\underline{x}=U\underline{v}$ for some $\underline{v}$

The vector $\tilde{\underline{x}}$ that solves this problem can be obtained in a particularly simple way: define the M by M matrix $T=(U^*U)^{-1/2}U^*RU(U^*U)^{-1/2}$; then $\tilde{\underline{x}}=U(U^*U)^{-1}\underline{e}$, where $\underline{e}$ is the normalized eigenvector of T associated with the smallest eigenvalue. We can now determine (r_0,z_0) by searching for the zero of the function $\underline{p}(r,z)^*\tilde{\underline{x}}$, or, having obtained T, we can calculate the <u>reduced maximum likelihood (RML) estimator</u>:

$$RML(r,z)=1/\,\underline{g}(r,z)^*T^{-1}\underline{g}(r,z)\ ,\ \underline{g}(r,z)=(U^*U)^{-1/2}U^*\underline{p}(r,z)\ . \qquad (10)$$

which is equivalent to a ML procedure on the mode-filtered R:

$$RML(r,z)=1/\,\underline{s}(r,z)^*[\underline{s}(r_0,z_0)\underline{s}(r_0,z_0)^*+Q]^{-1}\underline{s}(r,z) \qquad (11)$$

Reduced maximum likelihood method for the general case

What causes difficulty for the ML estimator in the normal mode case is that the modal noise resembles actual sources and that its matrix, UQU^+, has rank M but dimension N. The ML method and other high resolution methods expect the lowest eigenvalues to correspond to eigenvectors orthogonal to the signal component, but not to all potential signals as well. When this happens the ML is useless and the reduced ML, or something like it, required.

Let us suppose that the matrix R has rank M, but dimension N>M, so that it has the form $R=UTU^+$, where T is an M by M positive definite Hermitian matrix and U is some N by M matrix; in the normal mode case it consists of eigenfunction samples, but in general, we will not have a model for U. Assume, for now, that U is known and that T is data dependent. Writing $T=WW^+$, where W is an M by M matrix, we have $R=UWW^+U^+=VV^+$, for $V=UW$. Because R has rank M its inverse does not exist, so ML cannot be formed. A standard trick to create invertibility is to add a small positive quantity to the main diagonal of R: in the normal mode case above this was done through the εI term, but did not make ML usable. Instead, we consider replacing R^{-1} by $R^{\#}$, the pseudo-inverse of R, to obtain a reduced ML estimator.

The pseudo-inverse of R is $R^{\#}=V(V^+V)^{-2}V^+=U(U^+U)^{-1}T^{-1}(U^+U)^{-1}U^+$ and the reduced ML estimator is

$$RML(\theta)= 1/\underline{p}(\theta)^+R^{\#}\underline{p}(\theta) = 1/\underline{s}(\theta)^+T^{-1}\underline{s}(\theta) \ , \qquad (12)$$

where $\underline{s}(\theta)= (U^+U)^{-1}U^+\underline{p}(\theta)$; this agrees with the RML we obtained above in the normal mode case.

In general we may not have a model for U and may need to form $R^{\#}$ some other way. We can do that by taking $R=XLX^+$, $R^{\#}=XL^{\#}X^+$, where $L^{\#}$ is defined to be the diagonal matrix with entries λ_n^{-1}, provided λ_n is not too close to zero, and 0 otherwise. Using $R^{\#}$ in (12) gives the RML estimator in the general case.

Using the matrix structure to obtain environment parameters

In [ ] Buckingham and Jones discuss the use of the angular distribution of noise power received by a vertical array to estimate the critical grazing angle of the bottom, and hence the compressional sound speed in the bottom sediment. This is an example of an inverse problem, in which environmental parameters are estimated by comparing measured data with theory.

In another paper [ ] Buckingham considers the structure of the modal noise component UQU^*, for the case of a range-independent isovelocity channel with a low-loss bottom, where the noise energy is uniformly distributed immediately beneath the surface. The interesting point made in [ ], from our perspective, is that the matrix Q is essentially a multiple of the identity matrix. Knowing that the theory requires this and having measured $R=UQU^*$ in the noise-only case, we could estimate the matrix U: for each choice $\tilde{U}$ form $\tilde{Q}=(\tilde{U}^*\tilde{U})^{-1}\tilde{U}^*R\tilde{U}(\tilde{U}^*\tilde{U})^{-1}$; if we have chosen the correct $\tilde{U}=U$ then $\tilde{Q}=Q=\alpha I$, whereas for wrong choices of $\tilde{U}$ the matrix $\tilde{Q}$ need not be diagonal.

An interesting object of study would be those noise fields for which the matrix Q is diagonal. Is it generally true in normal mode environments or are such noises special cases? It is not a statistical phenomenon but, in the case considered in [ ] at least, follows from the nature of the eigenfunctions $U_m(z)$ and the manner in which the sources of independent noise energy are distributed spatially. The analysis in [ ] does invoke a "large M" approximation (p.1188, (10)), so leaves open the question of whether or not Q is diagonal for small M.

draft

On entropy criteria for solving inverse problems with positivity constraints

by

Charles L. Byrne and Lee K. Jones
Department of Mathematics
University of Lowell
Lowell, Massachusetts 01854

ABSTRACT

1. Introduction and Preliminaries

We are concerned here with suitable criteria to use for the reconstruction of positive functions from limited data. We shall assume that the integral of the function (call it R(x)) over its (compact) support is known, so that, by rescaling to have integral one, we may assume that the function to be recovered is a (measurable) probability density function. A general procedure for reconstruction is to select a prior estimate of R(x) (call it P(x)) and then to accept as the estimate of R(x) that data consistent density Q(x) that is closest to P(x), in some appropriate measure of distance. The distances $D(P,Q)=\int f(Q(x),P(x))dx$, where $f(y,z)$ is a suitably constrained function of $y,z>0$, provide a wide class of reconstruction procedures, which is the topic of this paper. A special case of such reconstruction is the "minimization of cross entropy" method (MCE).

The method of minimization of cross entropy (MCE), implicit in the work of Shannon [1], and advocated by numerous authors, including Jaynes [2,3], was proposed by Kullback [4], who called it the "principle of minimum directed divergence". The term "cross entropy" is due to I.J. Good [5]. The MCE method has been studied extensively by Shore and Johnson, who have derived the principle from axioms of consistent inference [6] and have used the resulting reconstruction method in speech processing and spectrum analysis [7].

Although the axiomatic derivation of the MCE method found in [6] is based on probability theory, it is finding application in the reconstruction of essentially non-probabilistic functions (which happen to have the mathematical properties of probability density functions), such as energy distribution as a function of bearing (array processing), x-ray attenuation functions (in tomography), and non-negative images (in optics). The properties that the reconstruction will exhibit in these cases are not easily predicted from the axioms of logical inference from which the method is derived. The basic problem is one of approximation; we wish to use the data and prior information to construct a function that will approximate the desired function in some appropriate sense. The MCE method employs a (non-symmetric) measure of distance between functions that obeys an orthogonality principle analogous to that associated with the metric of Hilbert space.

In [6] it was shown that this orthogonality principle serves to distinguish the MCE method from other distance minimization procedures obtained from the so-called Ali-Silvey-Csiszar distances. In this paper we consider the extent to which the orthogonality principle further distinguishes the MCE method from methods based on a much wider class of distances.

The problem is to reconstruct the non-negative measurable function $R(x) \geq 0$, defined on a compact set X within d-dimensional real Euclidean space, knowing only the linear functional values r_k, k=0,...,K, given by

$$r_k = \int R(x)\, g_k(x)\, dx\,, \qquad (1)$$

where the $g_k(x)$ are known linearly independent bounded measurable functions. We assume that $g_0(x)=1$ for all x, so that r_0= area of R(x), which we then take to be equal to one. We let $\mathcal{Q}$ be the collection of all measurable probability density functions supported on X for which (1) holds. A member of $\mathcal{Q}$ is called *admissible* if it is bounded above and away from zero on X. We assume that we have available a prior estimate of R(x) in the form of a measurable probability density function P(x), supported in X. To obtain our reconstruction (estimate) of R(x) we select that Q(x) in $\mathcal{Q}$ closest to P(x) in some appropriate sense.

To measure closeness we employ non-negative distances of the form

$$D(Q,P) = \int f(Q(x),P(x))\, dx\,, \qquad (2)$$

where f(y,z) is a suitably smooth kernel function. Associated with a particular distance is the following optimization problem:

Problem A: Find $\tilde{Q}$ in $\mathcal{Q}$ such that $D(\tilde{Q},P) \leq D(Q,P)$, for all Q in $\mathcal{Q}$.

For example, the MCE method employs the kernel f(y,z)= y log(y/z), and D(Q,P) becomes the *cross entropy* of Q, given P:

$$E(Q,P) = \int Q(x) \log[Q(x)/P(x)]\, dx\,; \qquad (3)$$

the MCE solution is that $Q_{MCE}(x)$ in $\mathcal{Q}$ for which (3) is minimized.

2. The orthogonality principle for the MCE method

It may be that there is no member of $\mathcal{Q}$ for which $E(Q,P)$ is finite; suppose the support of P is not all of X and $g_1(x)$ is 1 for those x in X, but outside the support of P, while 0 otherwise, and that $r_1=1$. However, if there is one Q in $\mathcal{Q}$ for which $E(Q,P)$ is finite the the MCE solution exists and has the form

$$Q_{MCE}(x) = P(x) \exp(a_0 g_0(x) + \ldots + a_K g_K(x)) , \qquad (4)$$

for x in N, and zero otherwise, where N is the union of the supports of all Q in $\mathcal{Q}$ for which $E(Q,P)$ is finite. A rigorous proof of this was first given by Csiszar [9] and a more restrictive sufficiency version was proved earlier by Kullback [10].

Assume now that the support of P equals X, so that the MCE solution exists and has support X. Let $\mathcal{T}$ be the set of all admissible densities having the form

$$T(x) = P(x) \exp(t_0 g_0(x) + \ldots + t_K g_K(x)) , \; x \text{ in } X. \qquad (5)$$

Since $\int T(x)\,dx = 1$ there are only K free parameters and $t_0 = t_0(t_1,\ldots,t_K)$ is a function of the other parameters; specifically,

$$t_0 = -\log \{ \int P(x) \exp(t_1 g_1(x) + \ldots + t_K g_K(x)) . \qquad (6)$$

The following theorem is the *orthogonality principle* of interest here:

Theorem: The choice $T(x)=Q_{MCE}(x)$ minimizes $E(R,T)$ over all T in $\mathcal{T}$.

proof: It is easily shown that $E(R,T)=E(Q_{MCE},T) + E(R,P) - E(Q_{MCE},P)$, so that $E(R,T)$ and $E(Q_{MCE},T)$ are simultaneously minimized. But it follows from elementary properties of the function $f(y,z)=y \log(y/z)$ that $E(Q_{MCE},Q_{MCE}) \leq E(Q_{MCE},T)$, with equality if and only if $T=Q_{MCE}$.

In this paper we are concerned with characterizing those distances $D(Q,P) = \int f(Q(x),P(x))\,dx$ for which the analogous orthogonality principle holds.

3. The generalized orthogonality principle

If P is an admissible density and the $\tilde{Q}(x)$ in $\mathcal{Q}$ minimizing D(Q,P) is also admissible then it can be shown that the Euler-Lagrange equation must be satisfied; that is, for all x we have

$$f_1(\tilde{Q}(x),P(x)) = a_0\, g_0(x) + \ldots + a_K\, g_K(x)\,, \qquad (7)$$

for some choice of constants $a_0,\ldots,a_K$, where $f_1(y,z)$ is the partial derivative of f with respect to the first variable.

We now generalize the orthogonality principle by defining the class $\mathcal{T}$ now to be all admissible densities T(x) that, for some constants $t_0, \ldots, t_K$, satisfy the equation

$$f_1(T(x),P(x)) = t_0\, g_0(x) + \ldots + t_K\, g_K(x)\,, \text{ for all } x. \qquad (8)$$

We then consider a second problem associated with the distance D:

Problem B: Minimize D(R,T) over all T in $\mathcal{T}$.

The distance measure D(Q,P) is then said to exhibit the *orthogonality principle* if whenever Problem A has unique solution $\tilde{Q}(x)$ then it is also the unique solution to Problem B.

Example 1. If $f(y,z) = y \log(y/z)$ then D(Q,P)=E(Q,P) and the orthogonality principle holds.

Example 2. If $f(y,z) = (y-z)^2/2$ then $\tilde{Q}(x) = P(x) + a_0\, g_0(x) + \ldots + a_K\, g_K(x)$ follows from (8); although the orthogonality principle holds there may be no positive member of $\mathcal{Q}$ of this form.

Example 3. If $f(y,z) = y\, e(z/y)$, where $e(t)$ is three times continuously differentiable and $d^2e/dt^2 > 0$, then the distance $D(Q,P)$ is in the smooth Ali-Silvey-Csiszar class; it was shown in [8] that only $E(Q,P)$ satisfies the orthogonality principle in this class.

mention after def of $\perp$

We begin now to extend this result of [8] to a wider class of distances associated with more general functions $f(y,z)$.

4. Regularity assumptions and admissible kernels

We consider a general distance measure $D(Q,P)=\int f(Q(x),P(x))\, dx$, and impose conditions on the function $f(y,z)$ in order that the distance D have reasonable metric properties.

Condition 1: $f_1(y,y) = c = \text{constant}$, for $y>0$;

we want $D(Q,P) \geq D(Q,Q)$ always and the Euler-Lagrange equation corresponding to minimizing $D(Q,P)$, subject only to $Q>0$ and $\int Q = 1$, is $f_1(P(x),P(x))=\text{constant}$.

Condition 2: $f(y,y) = 0$ for $y>0$;

we want $D(P(x),P(x))=0$ for all $P(x)$.

If we let $h(y)=f(y,y)=0$ then $dh/dy = 0$. But $dh/dy = f_1(y,y) + f_2(y,y)$ so it follows from Conditions 1 and 2 that

Condition 3: $f_2(y,y) = -c$, $y>0$.

Our fourth condition is that $f(y,z)$ be strictly convex in the first variable; that is,

Condition 4: $f_{11}(y,z)>0$, for all $y,z>0$;

we arrive at this condition by examining what happens to $D(Q,P)$ for Q

near $\tilde{Q}$. Let x_0 be a fixed point in X and let U be a neighborhood of x_0. Then let h(x) be any continuous function, supported on U , and not in the linear span of the functions $\chi_U\, g_k(x)$, k=0,...,K, the restrictions to U of the functions $g_k(x)$. Let j(x) be the projection of h(x) onto the span of the $\chi_U\, g_k(x)$ and let m(x)= h(x)-j(x); then $\int m(x)\, g_k(x)\, dx = 0$, k=0,...,K. The functions $\tilde{Q}(x) + \varepsilon m(x)$ are data consistent, for all $\varepsilon>0$. We want the minimum of $\int f(\tilde{Q}(x)+\varepsilon m(x),P(x))$ to occur at $\varepsilon=0$; the second derivative, with respect to ε, is $\int f_{11}(\tilde{Q}(x)+\varepsilon m(x),P(x))\, m(x)^2\, dx$. This must be positive at $\varepsilon=0$, for every neighborhood U, no matter how small, and for every point x_0; from our freedom to select the problem, hence the $\tilde{Q}(x)$ and P(x), it follows that Condition 4 must hold.

Summarizing, we say that f(y,z) is an *admissible kernel* if the following conditions hold: $f_1(y,y)=-f_2(y,y)=c$; $f(y,y)=0$, $f_{11}(y,z)>0$, for all y,z>0.

The four conditions above do not guarantee that the $\tilde{Q}(x)$ satisfying the Euler-Lagrange equation will be positive; the function f(y,z) is said to be *p-admissible* if the equation (7) always has a positive solution for $\tilde{Q}(x)$.

5. Necessary conditions for the orthogonality principle

The following Proposition (which we prove in the Appendix) is basic to our characterization of distances $D(Q,P)=\int f(Q(x),P(x))\, dx$ having the orthogonality principle:

Proposition 1: Suppose that f(y,z) is an admissible kernel such that, for every choice of R(x) , the $g_k(x)$ and the prior P(x), the orthogonality principle holds for $D(Q,P)=\int F(Q(x),P(x))dx$. Then $f_{211}(y,z)=0$ for all y,z>0.

From the equation $f_{211}(y,z)=0$ and our conditions it follows that f(y,z) has the form

$$f(y,z) = cy - yj(z) + J(y) - cz + zj(z) - J(z) , \qquad (9)$$

for some function strictly convex J(z), with dJ/dz= J'(z)=j(z). The strict convexity of J(z) implies that the function $\tilde{f}(y,z)$, given by

$$\tilde{f}(y,z) = -yj(z) + J(y) + zj(z) - J(z) , \qquad (10)$$

is positive. It also follows that if f(y,z) has form (9) then the necessary condition (7) becomes

$$f_1(\tilde{Q}(x),P(x)) = c + j(\tilde{Q}(x)) - j(P(x)) = 0. \qquad (11)$$

In order that f be p-admissible it is necessary and sufficient that the strictly increasing function j(y) map the positive reals onto the entire real line; we make this another condition:

Condition 5: The range of j(y), for y>0, is the entire real line.

Example 1: Let j(z)= log(z), so that J(z)= zlog(z) - z ; then condition 5 is satisfied. For c=1 we have f(y,z)= ylog(y/z), so the distance is cross entropy.

Example 2: Let j(z) = -1/z , so that J(z)= -log(z); then condition 5 is not satisfied. For c=0 we get f(y,z)=(y/z) - log(y/z). If P(x)= constant, for all x, then the resulting D(Q,P) is equivalent to negative Burg entropy, $-\int\log(Q(x))dx$. The Euler- Lagrange equation in this case leads to $\hat{Q}(x)= 1/(a_0+\ldots+a_K\, g_K(x))$, and the orthogonality principle says that the "maximum entropy method" (MEM) solution is the closest to R, among all density functions of its form, in the sense of minimizing the distance D(R,T)= $\int[(R/T) - \log(R/T)]$. We shall return to the MEM later.

6. Characterizing the MCE method

We have seen that (9) is necessary (along with our four conditions) for D(Q,P) to have the orthogonality principle generally. In fact the converse is also true:

Proposition 2: If f(y,z) satisfies (9) for all y,z>0 then D(Q,P) has the orthogonality principle.
proof: Let $\tilde{Q}(x)$ be the unique solution to Problem A. The Euler-Lagrange equation is now $j(\tilde{Q}(x))-j(P(x))= a_0+\ldots+a_K g_K(x)$ so the condition defining the class $\mathcal{T}$ is $j(T(x))-j(P(x))= t_0+\ldots+t_K g_K(x)$ for some constants $t_0,\ldots,t_K$. Since $\tilde{Q}$ is data consistent it follows that $D(R,T)=D(\tilde{Q},T)+D(R,P)-D(\tilde{Q},P)$, from which we conclude that $D(R,T)$ and $D(\tilde{Q},T)$ are simultaneously minimized. But $D(\tilde{Q},\tilde{Q})<D(\tilde{Q},T)$ unless $T=\tilde{Q}$.

We see from (9) that f(y,z) contains terms that involve only z; these terms, when they appear in D(Q,P), will involve only the prior P, and not the unknown Q. It would seem artificial if D(Q,P) has the orthogonality principle, but the distance obtained from D(Q,P) by omitting the terms that do not involve Q does not. In fact, we can characterize the MCE in precisely these terms; f(y,z)=ylog(y/z) is essentially the only function of the form (9) having no terms involving only the variable z.

Proposition 3: If f(y,z) has the form (9) and $zj(z) - J(z) - cz = 0$ for all z then D(Q,P)=jf(Q,P) is equivalent to E(Q,P).
proof. From the differential equation for J it follows easily that $J(z)=a(z\log(z)) + bz$, for some constants a and b.

Proposition 4: If D(Q,P)=D(P,Q) has the orthogonality principle then D(Q,P) is equivalent to least squares; that is, to the distance $(Q-P)^2$; hence there is no symmetric distance satisfying condition 5 that obeys the orthgonality principle.
proof: The symmetry condition and (9) imply that J(y) satisfies the following differential equation, for each fixed z:

$$J'(y)-(2/(y-z))J(y) = -(2/(y-z))J(z)-J'(z) \quad . \qquad (11)$$

It then follows that J(y) is quadratic in y.

7. The maximum entropy method of Burg and its extensions

In [ ] Burg considers the problem of reconstructing the positive power spectral density function $R(\omega)$, $|\omega| \le \pi$, from the finitely many values of its Fourier transform,

$$r_n = \int_{-\pi}^{\pi} R(\omega) \exp(-in\omega)\, d\omega/2\pi \quad , \quad |n| \le N \quad . \qquad (12)$$

The MEM he proposes adopts, as the estimate of $R(\omega)$, that positive function $Q(\omega) = Q_{MEM}(\omega)$ satisfying the constraints (12) for which the Burg entropy, $\int \log(Q(\omega))d\omega$, is maximized. The Euler-Lagrange equation says that $Q_{MEM}(\omega)$ must have the form

$$Q_{MEM}(\omega) = 1/\sum_{n=-N}^{N} b_n \exp(in\omega) \quad , \quad |\omega| \le \pi, \qquad (13)$$

where the b_n are chosen so as to satisfy (12).

It does not follow from (13) that there will be a positive solution. However, Burg proceeds by assuming that there is a positive solution of the form (13), and uses the Fejer-Riesz theorem ([ ], p. 231) to rewrite the solution in the form

$$Q_{MEM}(\omega) = a_0 / \left| \sum_{n=0}^{N} a_n \exp(in\omega) \right|^2 , \quad |\omega| \le \pi, \qquad (14)$$

where the polynomial $A(z) = a_0 + a_1 z + \ldots + a_N z^N$ has all its roots outside the unit circle. He then shows that the vector $\underline{a} = (a_0, \ldots, a_N)^T$ must satisfy the matrix equation $R\underline{a} = \underline{\delta}$, where R is the $N+1$ by $N+1$ matrix with entries $R_{m,n} = r_{n-m}$, $m,n = 1, \ldots, N+1$, and $\underline{\delta} = (1,0,\ldots,0)^T$. So far everything is based on the assumption that a positive solution exists.

But now Burg shows that, given the matrix R obtained from the data, the $\underline{a}=R^{-1}\underline{c}$ must be such that A(z) has all its roots outside the unit circle (that is, $\underline{a}$ has the *minimum phase property*), from which it follows that the $Q_{MEM}(\omega)$ in $(1\frac{5}{4})$ is data consistent. So he has succeeded in producing a positive, data consistent function. The question still remains: Does this $Q_{MEM}(\omega)$ actually maximize the Burg entropy, among the class of data consistent power spectral densities? .

That it does maximize the entropy follows from a consideration of the relationship between the entropy and the error of one-step prediction of a time series from knowledge of its infinite past (see Papoulis [ ], p.427).

We noted earlier that $f(y,z)=(y/z)-\log(y/z)$ gives the MEM, but the j(y) fails to have condition 5. This suggests that, in general, the MEM formalism may not lead to a positive solution; this is the case, for example, when R is defined on higher dimensional space, or when the functions g_k are not simply one-dimensional exponentials.

It also causes difficulty if, in the problem considered by Burg, the support of $R(\omega)$ is $[-\Omega,\Omega]$, instead of $[-\pi,\pi]$, where $0<\Omega<\pi$. The polynomial in $(1\frac{4}{4})$ need not factor as before; all that the Euler-Lagrange equation requires is that the solution have the form $(1\frac{4}{4})$ over $[-\Omega,\Omega]$, so the polynomial can be negative for some values of ω. There is no guarantee that the function of the form $(1\frac{4}{4})$ that is data consistent will be positive within $[-\Omega,\Omega]$.

Burg's maximum entropy method (MEM) for reconstructing a positive function from Fourier transform values is justified as a method for power spectrum estimation by appealing to the limiting expression for the multivariate entropy of time-domain samples of a Gaussian random process. As we have just seen, the maximization of $\int\log(Q(x))dx$, subject to the data constraints, corresponds to the use of $j(z)=-1/z$ in (9), hence has the orthogonality principle. The Euler-Lagrange equation shows that the MEM solution has the form $\tilde{Q}(x)= 1/\{a_0+\ldots+a_K\, g_K(x)\}$, with the a_k chosen so as to make $\tilde{Q}$ data consistent. From Proposition 2 we know that, whenever $\tilde{Q}(x)$ exists, it provides a minimum, among all densities of its form, for the distance $\int[(R/T)-\log(R/T)]\,dx$. Thus $\tilde{Q}(x)$ is close to R(x) in an appropriate sense and the MEM is justified without appeal to Gaussian processes.

5. Summary

We have considered reconstruction procedures based on the minimization of a distance $D(Q,P)=\int f(Q(x),P(x))dx$, subject to Q satisfying data constraints, where P(x) is a prior estimate of the positive function to be recovered. We have limited the discussion to those f(y,z) satisfying certain conditions chosen to endow D(Q,P) with properties suitable for approximation. The MCE method, based on the choice of $f(y,z)=y\log(y/z)$, obeys an orthogonality principle analogous to that associated with orthogonal linear projection in Hilbert space, and our objective here was to investigate the extent to which this orthogonality principle serves to characterize MCE among distances of the above form. Such a characterization would then provide a purely approximation theoretic justification for the MCE procedure.

We have shown that, in order for the distance $D(Q,P)=\int f(Q(x),P(x))dx$ to obey the orthogonality principle (as well as the other conditions) it is necessary that f(y,z) have the form

$$f(y,z)= cy - yj(z) + J(y) - cz + zj(z) - J(z), \quad y,z>0, \qquad (16)$$

where J(y) is a strictly convex function with derivative j(y), and c is some constant. Terms involving z only play no role when D(Q,P) is minimized as a function of Q; if those terms vanish identically then the distance D is equivalent to cross entropy.

Appendix 2: Proof of Proposition 1

If we formally differentiate the equation (7) defining the class $\mathcal{T}$, with respect to the variable t_k, for some k=1,...,K, we obtain

$$(\partial T/\partial t_k) f_{11}(T(x),P(x)) = \partial t_0/\partial t_k + g_k(x) \ . \qquad (17)$$

Fix x and consider (16) as a partial differential equation for the function $T=T(x;t_0,\ldots,t_K)$, viewed as a function of $t_0,\ldots,t_K$. We rewrite (16) as

$$\partial T/\partial t_k = [\partial t_0/\partial t_k + g_k]\, h(T)\,, \qquad (17)$$

where $h(y) = h(y;x) = 1/f_{11}(y,P(x))$. Let $G(y)=G(y;x)$ be the (increasing) anti-derivative of the function $h(y)$ having the "constant of integration" such that

$$G(\tilde{Q}(x)) = \sum_{k=0}^{K} a_k\, g_k(x)\,. \qquad (18)$$

Then the function

$$T(x;t_0,\ldots,t_K) = G^{-1}\Big(\sum_{k=0}^{K} t_k\, g_k(x)\Big) \qquad (19)$$

solves the partial differential equation (16), is equal to $\tilde{Q}(x)$ when the $t_k=a_k$ and is bounded above and away from 0, uniformly in x, for $\underline{t}=(t_1,\ldots,t_K)$ in a neighborhood of $\underline{a}=(a_1,\ldots,a_K)$.

For $\underline{t}$ in some neighborhood of $\underline{a}$ the function $T(x;t_0,\ldots,t_K)$ given by (19) will be a probability density function, hence the problem of minimizing $D(R,T)$ over T in $\mathcal{T}$ can be viewed as that of minimizing $D(R,T(\underline{t}))$ subject to the constraint $\int T(\underline{t}) = \int T(x;\underline{t})dx = 1$. Forming the Lagrangian we obtain

$$0 = \partial/\partial t_k \{ D(R,T(\underline{t})) - \lambda \int T(\underline{t}) \}\,\Big|_{\underline{t}=\underline{a}} \qquad (20)$$

or, differentiating under the integral,

$$0 = \int \{ [f_2(R(x),\tilde{Q}(x)) - \lambda][\partial t_0/\partial t_k + g_k(x)]\, /f_{11}(\tilde{Q}(x),P(x))\}\, dx\,. \qquad (21)$$

From $\int \overline{T}(\underline{t}) = 1$ it follows that (for $\underline{t}$ near $\underline{a}$, so that T is in $\mathcal{T}$)

$$0 = \partial/\partial t_k \{ \int T(\underline{t}) \} = \int [\partial t_0/\partial t_k + g_k]/f_{11}(T,P) , \qquad (2\overset{5}{I})$$

so that

$$0 = \int \{ [f_{\underline{2}}(R(x),\tilde{Q}(x))][(\partial t_0/\partial t_k)\big|_{\underline{t}=\underline{a}} + g_k(x)]/f_{11}(\tilde{Q}(x),P(x))\} \, dx. \qquad (2\overset{4}{J})$$

Equation (2J) holds if we replace R(x) with any other pdf that is data consistent. If x_0 is any point in X and U is a neighborhood of x_0 we can find a bounded measurable function $\tau(x)$, supported on U, and orthogonal to each of the g_k, so that $S(x) = R(x)+\varepsilon\tau(x)$ is a data consistent pdf, for $\varepsilon>0$ small enough; then $(2\overset{4}{J})$ holds with S(x) instead of R(x).

Replacing R(x) with S(x) and then differentiating twice with respect to ε gives

$$0 = \int \{f_{211}(S(x),\tilde{Q}(x))(\tau(x))^2[(\partial t_0/\partial t_k)\big|_{\underline{t}=\underline{a}} + g_k(x)]/f_{11}(\tilde{Q}(x),P(x))\} \, . \qquad (2\overset{5}{K})$$

The integral is over the neighborhood U, since $\tau(x)$ is zero off U.

Now let $z = \tilde{Q}(x_0)$ be some value in the range of $\tilde{Q}$, and $y>0$. By adding to R(x) a narrow Gaussian density centered at $x=x_0$ and then correcting by a linear combination of the g_k we can assume that $R(x_0)=y$. If, in the neighborhood U of x_0, the function $[(\partial t_0/\partial t_k)\big|_{\underline{t}=\underline{a}} + g_k(x)]$ does not change sign, then, by shrinking the neighborhood about $x=x_0$ and letting ε go to zero in S(x), we can conclude that $f_{211}(y,z)=0$.

Since we are interested in distances D(Q,P) that will be applied to a variety of problems we are free to tailor the problem to suit the technical assumptions needed above. In particular, we can assume that the x are one-dimensional, that the g_k are continuous in a neighborhood of x_0 and assume a given value only finitely many times. So, except perhaps for finitely many values of x_0, the above argument holds and $f_{211}(y,z)=0$ for $z=\tilde{Q}(x_0)$ and any $y>0$. But this must be true for all problems, so the conclusion then is that f_{211} must be identically zero, for $y,z>0$.

Title page only

Gabor representations and wavelets

John J. Benedetto*
Department of Mathematics
and Systems Research Center
University of Maryland

Hi Charlie

I hope this acknowledgement is OK and thanks again. Could you pass the other preprint onto Lee.

Best wishes for your research and "daining". Yours,

MD87-36-JJB
TR87-36

August 1987

*The author gratefully acknowledges support from NSF Grant DMS 8740113 and, during a one week visit to U. of Lowell, from ONR Grant N00014-87-K-0394.

Title page only

Approximation-Theoretic Derivation of Logarithmic Entropy Principles for Inverse Problems and Unique Extension of the Maximum Entropy Method to Incorporate Prior Knowledge

Lee K. Jones*

University of Lowell

*Research of this author supported in part by ONR contracts N00014-85-K-0402, N00014-85-C-2503, N00014-K-87-0394.

END

DATE

FILMED

6-1988

DTIC